2022 年第一期全国人工智能社会实验讲习班

教学简报

第 2 期 2022 年 7 月 27 日

讲习班第二天课程及案例研讨活动纪实

7 月 27 日，是 2022 年第一期全国人工智能社会实验讲习班开班的第二天。上午，中国人民大学社会学系赵延东教授围绕“人工智能社会实验研究的伦理规范和伦理制度建设问题”重点讲授了科技伦理规范、伦理审查以及审查机制建设等方面的内容。下午，清华大学智能社会治理研究院院长、全国人工智能社会实验专家组组长苏竣和清华大学公共管理学院博士后魏钰明主讲了“人工智能社会实验实施路径”课程。李雪梅对智能社会治理实验基地管理办法进行了介绍。在试验区专题研讨环节，科技部战略规划司综合处处长常歆识、副处长许谦参加了会议。清华大学公共管理学院教授王有强和山东财经大学教授刘培德分别对上海和济南国家新一代人工智能创新发展试验区的建设情况进行了介绍，科技部中国科学技术发展战略研究院原副院长杨起全进行了点评。



图 1　7 月 27 日杭州、成都、长春、武汉、枣庄、贵阳等教学点一览

讲习班课程纪实

第四讲和第五讲以“人工智能社会实验研究的伦理规范和伦理制度建设问题”为题，由中国人民大学社会学系教授赵延东主讲。赵延东指出人工智能技术对社会日益产生深刻的影响，也引发了人们对人工智能技术相关的风险和伦理问题的关注。针对人工智能的综合社会影响和社会治理问题开展的实验研究，不仅有“人工智能技术伦理”问题，更需关注“人工智能社会实验伦理”问题。第四讲和第五讲课程由华中科技大学公共管理学院教授徐晓林主持。



图 2 华中科技大学公共管理学院教授徐晓林主持第四讲和第五讲课程

赵延东介绍了科技伦理规范的基本原则，既要尊重自主性、保护个人隐私、减少不必要的风险，又要保证研究有利于人类社会或科学知识增长，同时还要注意程序公正、回报公正和分配公正。科技伦理属于基本科学规范，忽视科技伦理可能给研究对象、社会、环境带来不可逆的损失和伤害，还会破坏科技与社会的关系，导致公众对科学和新技术的怀疑，甚至会使科学技术发展自身可持续性受到威胁。

赵延东指出社会科学研究较多地以人为对象，更要注重科技伦理问题，但国内社会科学伦理问题仍然未受到充分重视。社会科学研究由于涉及的人、事、组织往往超过研究自身的范围，使得伦理责任划分比较困难，另外研究者对伦理原

则的理解和实践也存在着巨大差异。在进行社会科学研究时，要维持研究的客观性和完整性，尊重被研究对象的隐私和尊严，使他们的身体和精神不受伤害；研究内容涉及个人隐私时，需要征得被研究对象的同意。在研究各个阶段均需要考虑伦理因素，在研究开始的设计阶段，研究者需要审查研究团队组成的合理性、受试者的选择标准、研究方案是否存在利益冲突；在研究过程中，要注意隐私安全、参与者知情权的落实以及新出现的伦理问题；在研究结束后，要考虑如何保证研究成果和收益的公平分配以及数据安全保护等问题。



图 3 中国人民大学社会学系教授赵延东主讲第四讲和第五讲课程

赵延东着重介绍了人工智能社会实验的伦理规范问题。人工智能社会实验属于数据化和信息化时代的研究，其规模更大、影响范围更广、途径更多样，兼具技术和实验系统的双重叠加风险，现有伦理规范适应性大大降低，需要新的更具针对性、更细分的伦理规范。在一定程度上说，能否在人工智能社会实验项目中把好伦理关，是决定项目成败的关键所在。研究人员在未经参与者同意甚至在其没有意识到的情况下对其进行数据采集和实验的能力在迅速增强，且采集规模巨大，这些变化速度远超规范、规则和法律的修订速度，这也带来了诸多问题和伦理困境。人工智能社会实验的伦理风险既包括隐私保护问题、知情同意的困难，也涉及技术层面的算法偏见、数据权属和信息安全问题。因此在进行人工智能社会实验设计时，研究人员应该设法用侵害性、风险性更小的方案来替代实验，设法改进实验处理，以使其尽可能无害，并将实验参与者的数量减少到完成科学目标所需要的最小数目。在进行数据采集时，研究人员应对涉及伦理的项目数据采

取限制措施，保证数据访问仅限于可信任的人员，尽可能将数据标识并汇总。

赵延东强调，人工智能社会实验中要特别重视伦理审查问题。项目启动前要实施初始审查，重点要查阅研究方案、项目立项书、诚信承诺书、利益冲突说明等。项目启动后要进行持续审查，研究风险是否发生改变、风险评估是否依然合理、可能的获益相关预期是否合理等。不具备伦理审查条件的机构或个人，可申请委托有条件的机构伦理委员会进行伦理审查，或委托区域伦理委员会进行审查，委托审查要特别注意质量控制。社会科学的伦理审查要按照学术标准和社会标准进行，而不受行政和商业力量的干预，并根据不同研究采取相适应的审查形式，构建形成“政府牵头、协会主理、机构主营”的社会科学研究伦理审查机制。

赵延东介绍了“人工智能综合影响社会实验研究”项目在人工智能社会实验伦理治理方面进行的实践尝试。在项目实施之初，项目组专家便设计了“科研伦理自我审查报告”专用模板，要求各子课题根据模板要求，对课题研究工作进行伦理自查。其次项目组成立伦理审查咨询专家组，研究人工智能社会实验的伦理问题。最后项目组对各子课题研究人员开展伦理知识培训，以期提高人工智能社会实验的管理人员、科研人员和参与人员的科研伦理意识和伦理水平。

赵延东最后提出建设人工智能社会实验伦理治理规范的政策建议。首先，各级政府主管部门必须承担起人工智能社会实验伦理建设的管理责任；其次，高校、科研院所和企业等承担社会实验的单位应落实伦理建设的主体责任；再者，国家科技主管部门应加强对人工智能社会实验伦理问题的研究工作布置；最后，重大科技项目负责人应积极承担在研究工作中落实伦理治理的责任。讲授结束后，赵延东还就医学领域人工智能社会实验的伦理问题与听众进行了交流互动。

清华大学智能社会治理研究院院长苏竣和清华大学公共管理学院博士后魏钰明为全体学员带来题为“人工智能社会实验实施路径”的第六讲课程。该讲针对“作为基地和试验区建设核心任务的人工智能社会实验工作如何深入推进”这一核心问题，对社会实验的的三大主体、场景遴选、成果转化和常见问题等四个主要部分进行了介绍。第六讲课程环节由清华大学智能社会治理研究院副院长刘运辉主持。



图 4 清华大学智能社会治理研究院副院长刘运辉主持第六讲课程

苏竣强调人工智能社会实验既需要有理论支撑，也要有具体的、可行的实施方案。人工智能社会实验需要三大实施主体——应用主体、技术主体和研究主体分工合作、协同配合，地方政府需要在这个过程中承担组织建设的责任，为技术实施者提供服务，研究主体和技术主体要为社会实验提供决策支持。需要指出的是，社会实验也有一定的局限性，由于一些不确定因素的存在，人工智能社会实验的结果可能与现实存在着一些偏差，这也是难以规避的。



图 5 清华大学智能社会治理研究院院长苏竣主讲第六讲课程

魏钰明指出，人工智能社会实验是一项同时包含技术应用、科学研究和服务于治理的政治任务。围绕核心任务形成了社会实验的应用主体、技术主体和研究主体。三大主体彼此协同，互相合作，促进社会实验工作开展。应用主体为实验基地和试验区建设进行协调组织并提供资金支持，以政府部门为主；技术主体为实验基地建设提供智能技术服务与应用，以技术企业为主；研究主体为实验基地建设和社会实验开展提供学术支持和咨询指导，以高校、智库等科研机构为主。

魏钰明介绍了人工智能社会实验应用场景的遴选方式。场景遴选是阶段一（组织应用）和阶段二（科学测量）的过渡环节，也是社会实验是否“接地气”的关键一步。魏钰明将场景遴选的标准通俗易懂地概括为政府想建、企业能干、适合研究。目前，中央网信办、科技部等有关部委已先后建立了多家人工智能创新发展试验区和智能社会治理基地，搭建智能社会治理实验场景。魏钰明以宜老适老社会实验和黑土地保护社会实验为例，从宏观背景、实验方法、研究对象、研究问题和场景搭建等角度阐述了场景遴选的注意事项。



图 6 清华大学公共管理学院博士后魏钰明主讲第六讲课程

魏钰明从政府、企业和社会三个维度介绍了社会实验的成果转化。成果转化是人工智能社会实验阶段二（科学测量）和阶段三（综合反馈）之间的过渡环节。政府维度以政策报告为载体，要积极拓展人工智能社会实验研究成果的政策影响力，并将政策措施和经验做法转化为可供示范推广的经验成果；企业维度以技术

方案为载体，形成基于用户视角的技术优化路径和迭代方案；社会维度以标准化建设为载体，将人工智能社会实验作为智能社会治理标准化的重要方法。随后，魏钰明仍以宜老适老社会实验和黑土地保护社会实验两个典型项目为例介绍了成果转化的未来思路，为各基地的成果转化工作设计提供了必要的参考。

魏钰明认为，社会实验是高度复杂的系统工程，在不同环节都可能出现研究误差，应当引起研究者高度重视。魏钰明从微观的实验主体、中观的实验过程、宏观的实验设计三个层面介绍了可能存在的误差和问题，给出了伦理角度之外的避坑指南。在微观层面，罗森塔尔效应、霍桑效应、安慰剂效应等效应说明实验者或受试者因暴露在实验环境下，会产生由不同程度的心理和行为异化带来的误差；在中观层面，干预质量差异和样本不遵从等问题表明各类管理和执行因素导致实验整体效果会出现偏离理想状态的效果误差；在宏观层面，随机化偏误、溢出效应、均衡效应等效应告诉我们忽视某些关键要素会导致实验结果与社会现实存在一定距离。魏钰明也相应介绍了缓解或解决部分问题的方式方法，并建议加强实验队伍的标准化建设，方能保证实验的准确性和科学性。



图 7 李雪梅同志介绍国家智能社会治理实验基地管理办法

李雪梅同志依次从总则、申报程序、基地建设、指导与评估、附则向各位学员介绍了《国家智能社会治理实验基地管理办法（暂行）》（以下简称“管理办法”）。管理办法是在深入开展人工智能社会实验的背景下，由国家多部委联合

印发的指导性文件，适用于包括综合基地和特色基地在内的所有国家智能社会治理实验基地。

李雪梅建议，各基地应高度重视实验基地的建设工作，认真学习领会管理办法，保障资源投入和部门协同。基地应当广泛邀请专家为基地建设提供咨询指导。建立智能社会治理实验技术伦理和科研伦理审查机制。此外，基地应积极利用智能社会治理网的平台把本基地工作经验进行宣传推广。李雪梅从建设主体、实施保障、实施方案、工作进度、社会效益等五个方面介绍了基地评估检查的相关考核重点。建议各基地在此基础上进行自评估，以评促建，促进基地的社会实验开展。

课后，魏钰明和李雪梅针对在场听众关心的问题进行了解答。国家人口健康科学数据中心副主任、专家组成员尹岭，中央网信办信息化发展局转型发展处干部欧阳曾思，北京师范大学政府管理学院副教授郭跃就如何解决社会研究中的研究误差等问题进行了交流发言。尹岭建议管理办法作用客体应为项目而非试验区，并对社会实验基地进行适宜奖惩；欧阳曾思提问如何看待罗森塔尔效应与伦理原则之间可能存在的矛盾，魏钰明介绍了目前可以采取的实验设计方法；郭跃与魏钰明围绕人工智能社会实验如何真正做到“以人为本”进行了探讨。

试验区专题研讨

根据讲习班的教学安排，7 月 27 日下午组织了第一场专题研讨。科技部战略规划司综合处处长常歆识、副处长许谦参加了试验区案例研讨。清华大学公共管理学院教授王有强和山东财经大学教授刘培德分别对上海和济南国家新一代人工智能创新发展试验区的社会实验实施情况进行了介绍，科技部中国科学技术发展战略研究院原副院长杨起全进行了点评。研讨环节由清华大学智库中心副主任汝鹏主持。



图 8 清华大学智库中心副主任汝鹏主持试验区专题研讨

王有强结合上海试验区的情况进行了分享。王有强从实施概况、分析框架、指标体系、工作机制、场景示例、宣传外联六个方面回答了“如何设计”和“如何实施”两个核心问题。王有强首先介绍了上海人工智能社会实验实施概况的三层次。第一层级是实验基础路径，基于已有实践概况原创理论、基于原创理论建构可操作的分析框架、形成可跟踪的指标体系，并利用数据平台进行信息收集、整理和分析。第二层级为社会实验的具体应用场景，即人脸识别、“随申码”应用拓展、医疗诊断辅助决策支持、老人智能服务应用、临床医疗数据治理；第三层级为社会实验的目标展望。

王有强认为，成功的社会实验需要实现四方面的社会效益：学术研究、政策

研究、人才培养、实践创新。王有强随后介绍了可用于研究人工智能技术社会应用综合影响及其影响因素和机理的理论框架——系统模型 SMART。该模型由五个要素共同集合组成，以人工智能技术（Technology）为出发点，在观测应用结果（Result）后进行数据分析，最后分别从需求者（Application）、供给者（Management）和政府（Government）的角度评估智能技术的影响。在这一模型基础上，构建了“135”元指标体系，关注技术对三个不同主体的实际影响。

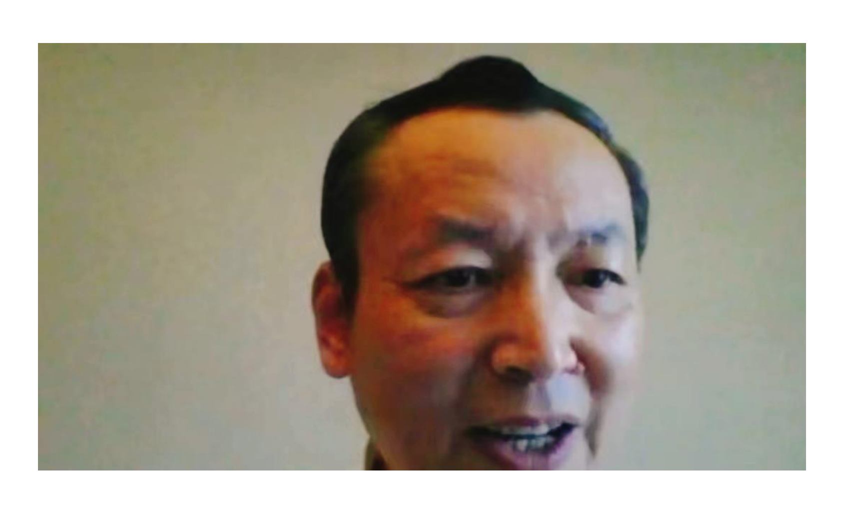


图 9 王有强介绍上海市人工智能社会实验实施情况

王有强以医疗诊断辅助决策支持系统为例介绍了上海市推进人工智能社会实验的具体实施方案。这一实验按照实验准备、测量指标、测评分析、数据汇总、研究成果五个环节的工作机制展开。该项目实施于上海复旦大学附属儿科医院，利用人工智能技术建立儿科门诊临床决策支持系统，通过智能诊断提示，减少错诊漏诊，减轻医生的工作负担。从需求、供给和政府方面设计了一系列的调研问题和分析指标，借助人工智能技术强大的算力和医院平台提供的海量数据。项目可以为医疗服务提供强有力的决策支持，也有助于推广智慧医疗，助力大健康产业可持续发展。

科技部中国科学技术发展战略研究院原副院长杨起全进行评论。杨起全认为上海市的人工智能社会实验研究方案完整详实、系统性强，做到了理论和实践相结合，为各基地开展 AI 社会实验研究提供了很好的借鉴。方案聚焦智能技术的社会问题，注重社会实验的方法设计，能够发掘科学技术的经济价值和社会价值。此外，通过使用科学的社会实验方法解构工具，形成了理论、框架、指标的完整系统，成功把泛义性的概念转化为可测量的指标，抓住了社会实验方法的精髓。

在研究建议方面，杨起全指出，各基地现有的研究在微观层面细致详实，但在宏观层面比较欠缺，局限于一城一地一事。建议从治国理政的高度发现问题，深入思考，由现象深入本质，充分发现规律；再者，应当更加关注人工智能技术介入长期化所带来的未来影响，根据近期变化对未来变化做出合理的科学预见；最后，应当关注标准化研究，形成可以推而广之的有效经验。



图 10 山东财经大学教授刘培德代表济南试验区进行汇报

刘培德代表济南试验区进行了汇报。刘培德分享了济南市人工智能社会实验的背景和总体思路。济南市以技术攻关为牵引，行业领域人工智能示范工程和典型场景打造为依托，选择实验组和对照组，开展人工智能社会实验，全力构建人工智能社会实验和人工智能重大示范工程相结合的“1+4+N”济南模式。

刘培德主要介绍了济南试验区的三个典型案例。案例一为“基于物联网的社区安全管理系统”。该案例遴选济南高新开发区 12 个社区居民，研究居民公共道德建立与和谐社区构建之间的路径和效应。此外，研究者重点关注老人轨迹追踪系统应用对和谐社区构建的影响。案例二为“远程智能医疗诊断平台”。该案例主要围绕“AI 技术对医生工作绩效以及医生与患者满意度构建作用路径与效应模型”的议题，检验了智能诊断对患者的医生信任和服务满意的影响机制。案例三为“复杂场景下金属零部件加工与缺陷智能检测系统”。该案例主要探索智能制造场景下 AI 应用对生产线工人心理及行为的影响路径及作用机理。



图 11 科技部中国科学技术发展战略研究院原副院长杨起全对案例进行点评

杨起全对济南试验区的分享进行了点评。杨起全认为，济南市人工智能社会实验场景类型和社会实验内容丰富，并且注重理论指导，形成了完整的研究假设和明确的实验问题，遵循着规范和扎实的学术研究范式。此外，其研究流程和核心概念的解析非常清晰，值得其他各基地借鉴学习。针对汇报，杨起全对济南市和其他各基地提出了几点建议：首先，关注敏感领域社会实验本身的伦理要求和伦理审查的具体实施问题，在执行过程中也要恪守伦理规范的要求；其次，要注意数据获取方式的科学性和数据分析的方式方法；最后，关注场景建设的不同过程对于人工智能社会实验的社会影响产生的作用。在互动交流环节，人工智能与数字经济广东省实验室（广州）战略研究中心副主管林韬杰分享了广州市开展人工智能社会实验的经验。

编辑：谢其军、徐诚、胡志明、蒋佳辰　　审定：苏竣、汝鹏

报送：中央网信办信息化发展局、教育部科学技术与信息化司、科技部战略规划司、民政部基层政权建设和社区治理司、民政部养老服务司、生态环境部信息中心、国家卫生健康委规划发展与信息化司、市场监管总局标准技术管理司、国家体育总局体育信息中心

抄送：各国家智能社会治理实验基地和国家新一代人工智能创新发展试验区

清华大学智能社会治理研究院、清华大学科教政策研究中心编印

电话： 010-62795573　　传真： 010-62795573